

Inapproximability of the Smallest Superpolyomino Problem

Andrew Winslow*
awinslow@cs.tufts.edu

Abstract

We consider the *smallest superpolyomino problem*: given a set of colored polyominoes, find the smallest polyomino containing each input polyomino as a subshape. This problem is shown to be NP-hard, even when restricted to a set of polyominoes using a single common color. Moreover, for sets of polyominoes using two or more colors, the problem is shown to be NP-hard to approximate within a $O(n^{1/3-\varepsilon})$ -factor for any $\varepsilon > 0$.

1 Introduction

The smallest superpolyomino problem restricted to one dimension yields the well-known *smallest superstring problem*: given a set of strings (i.e. $n \times 1$ polyominoes), find the smallest string containing each string in the set as a substring. This problem is known to be NP-hard and admits a simple greedy 3-approximation [1]. The approximation algorithm leads to a straightforward $O(\log^3 n)$ -approximation algorithm [2] for the minimum-size context-free grammar encoding a string, known as the *smallest grammar problem*.

In attempting to extend this idea to solve a higher-dimensional version of the smallest grammar problem, we consider the problem of finding the smallest superpolyomino for a given set of input polyominoes. We show that an approximation algorithm for the smallest grammar problem in two dimensions is unlikely to exist by proving that the smallest superpolyomino problem is not approximable within any $O(n^{1/3-\varepsilon})$ -factor for any $\varepsilon > 0$.

Moreover, this result remains true when constrained to sets of polyominoes using only two colors. We also show that the problem remains NP-hard even when constrained to sets of polyominoes using a single color. In this case, geometry alone is responsible for the problem difficulty, as the one-dimensional version can be solved in linear time with a single pass.

For the inapproximability result, we use a reduction from chromatic number of an arbitrary graph. Given a graph $G = (V, E)$, the reduction uses polyominoes each representing a vertex in V , with two polyominoes able to compactly

*Department of Computer Science, Tufts University. Research supported in part by NSF grants CCF-0830734 and CBET-0941538.

stack atop each other if and only if the two corresponding vertices are independent in G . Forming a smallest superpolyomino then is equivalent to finding a minimum coloring of the vertices of V , i.e. G 's chromatic number, shown to be NP-hard to approximate within any $(n^{1-\varepsilon})$ -factor by Zuckerman [4].

For the one-color case, a reduction from set cover is used. A small polyomino is created for each element of the universe, and a single large polyomino encoding the sets is used as a base for the final superpolyomino. Each element polyomino fits inside the set polyomino at locations corresponding to sets containing the element, with the exception of a single missing cell. A small superpolyomino then is obtained by placing the element polyominoes at a small number of locations inside the set polyomino, equivalent to covering the universe with a small number of sets.

2 Definitions

A polyomino $P = (S, C)$ is defined by a connected set of points on the square lattice (called *cells*) containing $(0, 0)$, and a coloring $C : S \rightarrow \sigma$ mapping the cells of P to a set of colors σ , e.g. cell $(3, 1)$ is red, cell $(3, 2)$ is gray, etc. As shorthand, we denote the color of the cell $C((x, y))$ as $P(x, y)$, and $|P|$ as the number of cells in P , i.e. the *size* of P . A *translation* of P is a polyomino $P' = (S', L')$ and coordinate offset (δ_x, δ_y) such that $S' = \{(x + \delta_x, y + \delta_y) \mid (x, y) \in S\}$ and $C'((x + \delta_x, y + \delta_y)) = C(x, y)$.

Two polyominoes $P_u = (S_u, C_u)$ and $P_v = (S_v, C_v)$ at some translation (δ_x, δ_y) are *compatible* if for each (x, y) , either $P_v(x, y)$ or $P_u(x, y)$ is empty or $P_v(x, y) = P_u(x + \delta_x, y + \delta_y)$. Similarly, a polyomino $P = (S, C)$ is a *superpolyomino* of $P' = (S', C')$ if there exists a translation (δ_x, δ_y) of P' such that for each (x, y) , either $(x + \delta_x, y + \delta_y) \notin S'$ or $P'(x + \delta_x, y + \delta_y) = P(x, y)$, i.e. there is a translation of P' such that P' is compatible with P and lies entirely in P .

3 Reduction

Given a graph $G = (V, E)$, each vertex $v \in V$ is converted into a polyomino $P_v = (S_v, C_v)$ that encodes v and the neighbors of v in G (see Figure 1). Each P_v is a rectangular $2|V| \times |V|$ polyomino with up to $|V| - 1$ single squares removed and lower-left corner at $(0, 0)$. The four corners of all P_v have a common set of four colors: green, blue, purple, and orange. Cells at locations $\{(2i + 1, 1) \mid 0 \leq i < |V|\}$ are colored black if $v_i = v$, r if $(v, v_i) \in E$, or are empty locations if v_i is not v or a neighbor of v . All remaining cells have a common gray color.

Consider how two polyominoes P_u and P_v can overlap depending upon the relationship of u and v . Because of the four distinct corner colors, P_u and P_v can only overlap when these four locations in P_v are translated to the same locations in P_u . In this translation, the cells at location $(2i + 1, 1)$ in P_u and P_v are compatible exactly when $(u, v) \notin E$, i.e. u and v are not neighbors. All other cells are colored gray and are thus compatible.

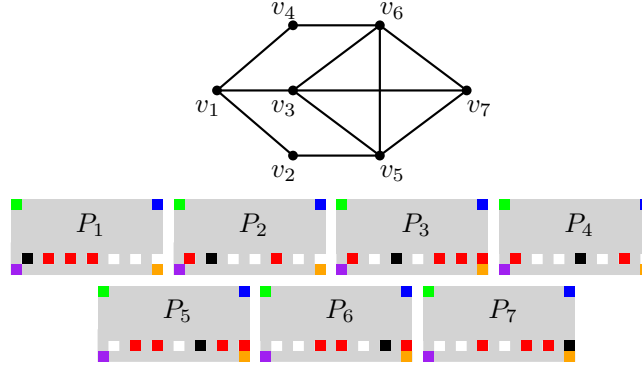


Figure 1: An example of the set of polyominoes generated from an input graph by the reduction.

The superpolyomino formed by a pair of compatible P_u and P_v in this translation has the common set of four colored corner cells and many gray cells, and has two black cells and a number of red cells corresponding to the combined neighborhoods of u and v . Then by induction, any set of polyominoes can overlap if and only if they form an independent set. Moreover, if they overlap, they overlap using a set of translations in which the four corners of all polyominoes are placed at four common locations.

Because the polyominoes can only overlap in this constrained way, any superpolyomino of the polyominoes $\{P_v \mid v \in V\}$ consists of a number of *decks* of superimposed polyominoes corresponding to independent sets of vertices in G arranged disjointly to form a single connected polyomino (see Figure 2).

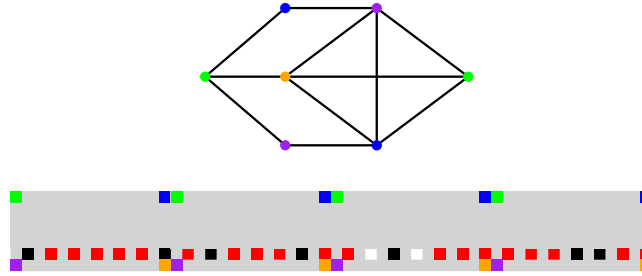


Figure 2: An example of a corresponding 4-deck superpolyomino and 4-colored graph.

Recall that each P_v is a $2|V| \times |V|$ rectangle with $|V|$ cells colored black, red, or are not present. The size of P_v is then between $2|V|^2 - |V| + 1$ and $2|V|^2$ depending upon the number of neighbors of v , and each deck of polyominoes also has size in this range.

Lemma 3.1 *For a graph $G = (V, E)$, there exists a superpolyomino of size at most $2k|V|^2$ for polyominoes $\{P_v \mid v \in V\}$ if and only if the vertices of V can be k -colored.*

Proof First, consider extreme sizes of superpolyominoes consisting of k and $k - 1$ decks. For any V and k with $1 \leq k \leq |V|$:

$$(k - 1)(2|V|^2) = 2k|V|^2 - 2|V|^2 < 2k|V|^2 - k|V| = k(2|V|^2 - |V|)$$

That is, the size of any superpolyomino of $k - 1$ decks is smaller than the size of any superpolyomino of k decks.

We now prove both implications of the lemma. First, assume the superpolyomino of size at most $2k|V|^2$ exists. Then the superpolyomino must consist of at most k decks. Each deck is the superposition of a set of polyominoes forming an independent set, so G can be k -colored.

Next, assume that G can be k -colored. Then the polyominoes $\{P_v \mid v \in V\}$ can be translated to form k decks, one for each color, each with size at most $2|V|^2$. Placing these decks adjacent to each other yields a superpolyomino of size at most $2k|V|^2$. \square

Note that only $|V|$ cells of each P_v are distinct and depend on v , while the other $2|V|^2 - |V|$ cells are identical for all P_v . The extra cells are needed for the first inequality in Lemma 3.1, and they effectively “drown out” variation in the size of each deck due to missing cells.

Theorem 3.2 *The smallest superpolyomino problem is NP-hard to approximate within a factor of $O(n^{1/3-\varepsilon})$ for any $\varepsilon > 0$.*

Proof Consider the smallest superpolyomino problem for the polyominoes generated from a graph $G = (V, E)$ with chromatic number k . There are $|V|$ of these polyominoes, each of size $\Theta(|V|^2)$, so the polyominoes have total size $n = \Theta(|V|^3)$. By Lemma 3.1, a superpolyomino of size between $(2|V|^2 - |V|)k'$ and $2|V|^2k'$ exists if and only if there exists a k' -coloring of G . Then by Zuckerman [4], finding a superpolyomino such that $(2|V|^2 - |V|)k'/(2|V|^2)k = O(|V|^{1-\varepsilon}) = \Theta(n^{1/3-\varepsilon})$ is NP-hard.

4 Extensions

The result of Theorem 3.2 also holds when constrained to sets of polyominoes using at most two colors by converting each cell into a unique 8×8 *macrocell* (seen in Figure 3). The cells on the boundary of the macrocell are colored black and the interior cells are colored gray, except for the three cells (2, 2), (3, 3), and (4, 4). These cells are used to encode a binary value between 0 and 7 corresponding to the color of the cell in the original polyomino, with gray indicating a 0 and black a 1.

Because all polyominoes are simply scaled by a factor of $8 \cdot 8 = 64$, following the proof of Lemma 3.1 gives a similar result, but with a superpolyomino size

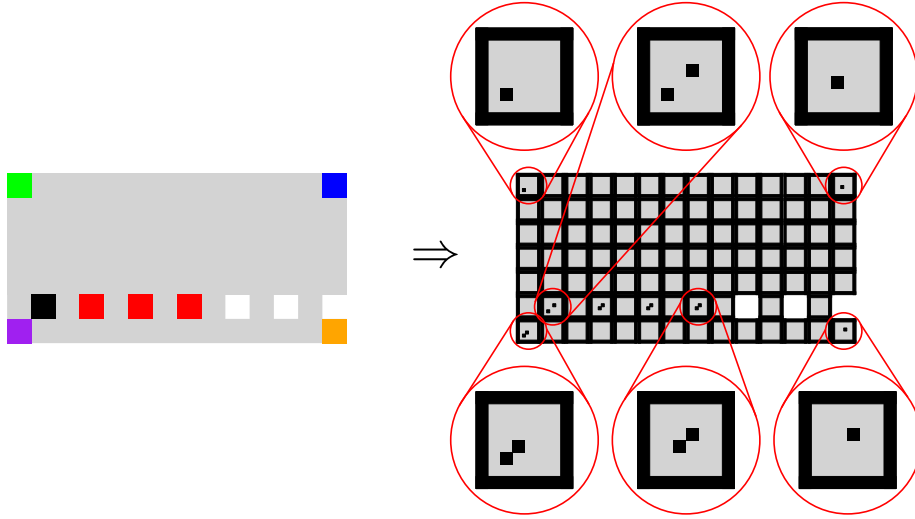


Figure 3: Converting a reduction polyomino (left) to a two-color reduction polyomino (right).

of $128k|V|^2$ instead of $2k|V|^2$. Because each polyomino still has size $\Theta(|V|^2)$, the same inapproximability ratio as Theorem 3.2 holds:

Theorem 4.1 *The smallest superpolyomino problem for two-color polyomino sets is NP-hard to approximate within a factor of $O(n^{1/3-\epsilon})$ for any $\epsilon > 0$.*

We now consider the smallest superpolyomino problem for one-color polyomino sets. A reduction from set cover is done using families of polyominoes as seen in Figures 4 and 5. Given a universe $U = \{1, 2, \dots, n\}$ and set of sets $S = \{S_1, \dots, S_m\}$ with $\bigcup_{1 \leq i \leq m} S_i = U$, two types of polyominoes are created. First, a set of *element polyominoes* $\{P_1, P_2, \dots, P_m\}$ is created, each consisting of a $(n+1) \times (n+1)$ *base*, a $1 \times 2n$ *flagpole*, and a $n \times 1$ *flag*. The lower-left corner of the three components for polyomino i are placed at $(0, 0)$, $(0, n)$, and $(1, n+2i)$, respectively.

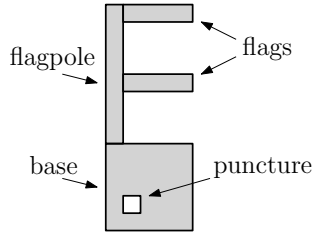


Figure 4: The components of element polyominoes and set gadgets used in the one-color reduction.

Second, a *set polyomino* \bar{P} is created, consisting of a set of *punctured bases* (a base with cell (1,1) missing) arranged horizontally and attached by single cells, each with a flagpole and some number of flags (called a *set gadget*). Each flagpole corresponds to a set $S_i = \{e_1, e_2, \dots, e_l\}$, with flags placed at location $((n+2)(i-1)+1, n+2e_j)$ for $1 \leq j \leq l$.



Figure 5: The set of polyominoes produced from the reduction from minimum set cover to smallest superpolyomino for the set $\{\{1, 2\}, \{1, 4\}, \{2, 3, 4\}, \{2, 4\}\}$.

Consider forming a superpolyomino of the set polyomino \bar{P} and some element polyomino P_i corresponding to an element $i \in U$. Define a translation of P_i that places the lower-left corner of its base at the lower-left corner of a punctured base of \bar{P} as an *alignment* of the bases. Aligning the base of P_i with a punctured base of \bar{P} corresponding to a set containing i results in a superpolyomino of size $|\bar{P}| + 1$. If the base of P is not aligned with a punctured base of \bar{P} , then the resulting superpolyomino is at least as large as some superpolyomino of \bar{P} and *all* element polyominoes:

Lemma 4.2 *Any superpolyomino of the set polyomino \bar{P} and some element polyomino P_i in which the base of P_i is not aligned with the punctured base of a set gadget not corresponding to a set containing i has size at least $|\bar{P}| + n$.*

Proof First, if the base of P_i is aligned with the punctured base of a set gadget that does not correspond to a set containing i , then the cells of the flag of P_i do not overlap with the cells of \bar{P} . The flag has n cells, and so the superpolyomino has at least $|\bar{P}| + n$.

Next, if the base of P_i is *not* aligned with any punctured base in \bar{P} , then it must be horizontally or vertically misaligned (or both). If the base is horizontally misaligned, then some column that does not contain any punctured base of \bar{P} contains cells from the base of P_i . Such a column then must contain at least $n+1$ cells of P_i and at most 1 cell of \bar{P} , so the superpolyomino has size at least $|\bar{P}| + n$.

If the base is not horizontally misaligned but is vertically misaligned, then some row containing a row of the base of P has at most 1 cell of \bar{P} . This row is either: 1. a row entirely below or above the set gadget P is aligned with or 2. a row that contains a cell of the flagpole but no flag. In either case, the superpolyomino has at least $n+1$ cells of P and at most 1 cell of \bar{P} , so the superpolyomino has size at least $|\bar{P}| + n$.

Intuitively, Lemma 4.2 implies that “playing by the rules” and aligning the base of each element polyomino with some punctured base of the set polyomino \bar{P} is always better than “cheating” by not aligning some element polyomino with any punctured base. So finding a minimum superpolyomino is equivalent to deciding which punctured base each element polyomino’s base should be aligned with, with a unit cost for each punctured base “patched” by an element polyomino’s aligned base (see Figure 6).



Figure 6: The smallest superpolyomino of the polyominoes in Figure 5, corresponding to the set cover $\{S_1, S_3\}$.

Theorem 4.3 *The smallest superpolyomino problem for one-color polyomino sets is NP-hard.*

Proof By Lemma 4.2, the smallest superpolyomino of any set of polyominoes generated by the reduction is one resulting from aligning the base of each element polyomino P_i base with the punctured base of a set gadget of \bar{P} corresponding to a set containing i . So a superpolyomino of size $|\bar{P}| + k$ exists if and only if there exists a set cover of size k . Set cover is known and loved by many, and is also NP-hard [3].

References

- [1] A. Blum, T. Jiang, M. Li, J. Tromp, M. Yannakakis, Linear approximation of shortest superstrings, *Journal of the ACM*, 41(4) (1994), 630–647.
- [2] M. Charikar, E. Lehman, A. Lehman, D. Liu, R. Panigrahy, M. Prabhakaran, A. Sahai, a. shelat, The smallest grammar problem, *IEEE Transactions on Information Theory*, 51(7) (2005), 2554–2576.
- [3] R. M. Karp, Reducibility among combinatorial problems, *Complexity of Computer Computations*, Plenum, New York, 1972, 85–103.
- [4] D. Zuckerman, Linear degree extractors and the inapproximability of max clique and chromatic number, *Theory of Computation*, 3 (2007), 103–128.